

Using *SocialAction* to Uncover the Catalano & Vidro's Social Structure

Adam Perer

Human-Computer Interaction Lab & Department of Computer Science, University of Maryland

ABSTRACT

This document characterizes the changes in the Catalano/Vidro social structure over the ten day period for the VAST Challenge 2008's Mini-Challenge 3.

KEYWORDS: VAST Challenge, SocialAction.

1 THE INITIAL COMPLEXITY

SocialAction, a tool that integrates statistics and visualization to improve exploratory data analysis [CITE], was the only tool used for this analysis. This document describes an overview of the analysis performed with *SocialAction* on the VAST Challenge data set. A corresponding video also highlights the discoveries. (<http://www.cs.umd.edu/~adamp/vast2008>)

The data from the cell phone logs was used to infer a social network. A node was created for each unique cell phone, and an edge was created for each phone call. This results in multiple edges between nodes if multiple phone calls between the two parties existed. This results in a complex social network as shown below, when all ten days are presented in Figure 1.

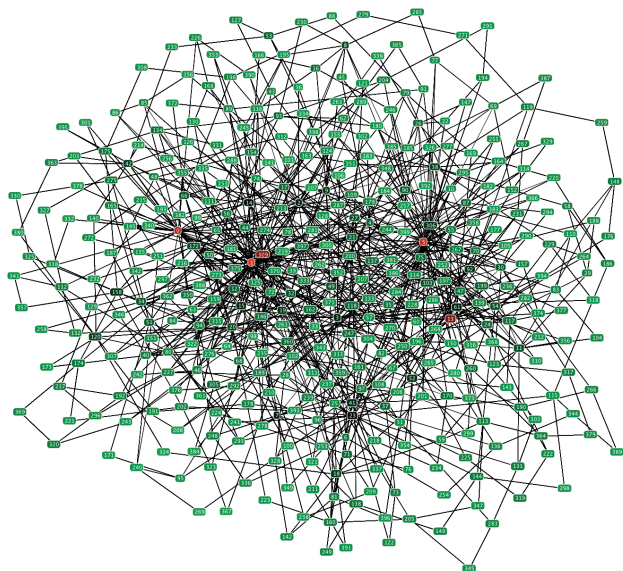


Figure 1. A network visualization of the entire cell phone log data set. Each node is a unique cell phone, and each edge is a call between cell phones. There are 400 nodes and 9834 edges when all 10 days worth of activities are loaded.

The social network visualization was generated using a force-directed layout. The nodes are colored according to their degree centrality on a red-black-green spectrum. Red nodes have the most neighbors, black nodes have an average number, and green nodes have fewer neighbors. *SocialAction* also provides a tabular, sortable, color-coded ranking list that is coordinated with the

visualization (refer to the video for a demonstration). Very little information seems to be evident from this chaotic network visualization. However, it is clear that there are a few individuals who are much more active with their cell phones than the rest of the network.

Since the mini-challenge was concerned with how the social structure changes over 10 days, we can also use *SocialAction* to display a timeline of the networks evolution over time. This results in the following interactive visualization displayed in Figure 2.

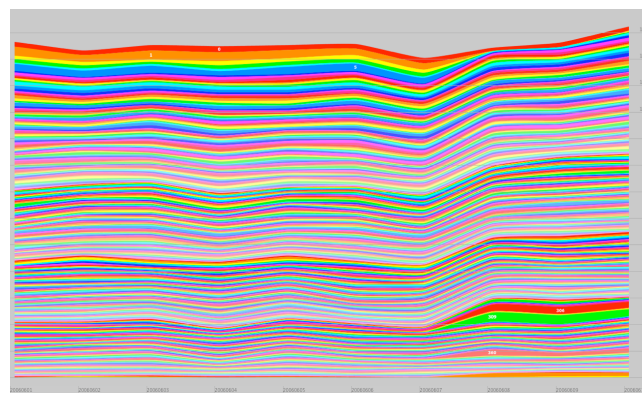


Figure 2. The timeline visualization of the 400 nodes over 10 days. Each stack represents a node, and the thickness of each stack on a given day represents its degree (the number of phone calls made and received). The visualization is interactive, as shown in the corresponding video.

This visualization is a stacked histogram, similar to NameVoyager. Each node in the network is assigned a stack, with a unique color. The thickness of each stack is according to the node's degree (the greater the number of phone calls initiated/received, the greater the thickness) at the time on the horizontal axis. Each day's activities have been binned into 10 daily cumulative value representing June 1-10, 2006.

Similar to the network visualization, this visualization is quite complex and hard to interpret. However, it is clear that there are a few individuals very active until June 8, and other individuals active beginning June 8. But who are these people?

2 IDENTIFYING THE CATALANOS AND VIDROS

VAST Intelligence suggests that there is medium confidence that Ferdinando Catalano is the node identified as 200. For the rest of my analysis, I assume this information is true. Since there were other clues about Ferdinando's close contacts, I decided to examine his immediate neighbors. Using *SocialAction*, I extracted the ego-network of Ferdinando and his neighbors. This ego-network is shown in Figure 3.

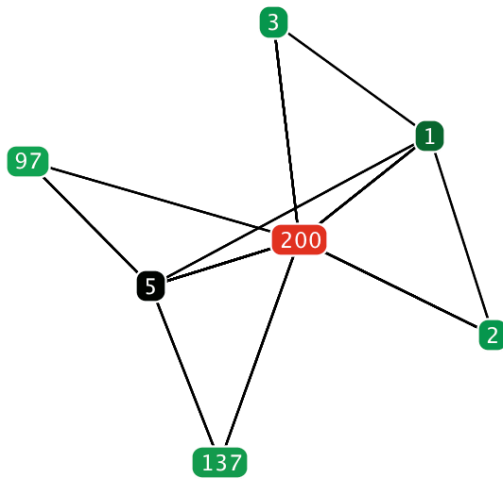


Figure 3. The ego-network of the node assumed to be Ferdinando Catalano (Node 200). This subnetwork shows Ferdinando's connections and their connections between each other. The nodes are colored according to their degree (red high, black average, green low).

Among Ferdinando's contacts, Ferdinando shares the most phone calls with Node 5, colored black. Due to this phenomenon, I assume that Node 5 is his brother Estaban, in accordance with statements from VAST intelligence ("We believe Ferdinando would call brother Estaban most frequently").

The next step was tracking down the Vidro brothers. VAST intelligence claims "David Vidro coordinates high-level Paraiso activities and communications". I therefore assume that he would be involved in the most calls among the people in the Paraiso movement. To approximate the Paraiso social network, I use *SocialAction* to extract another subnetwork: Ferdinando's contacts, and his contact's contacts. This subnetwork is shown in Figure 4.

The most active node in this subgraph is Node 1. Since this node has a structure similar to a coordinator, this node is assumed to be David Vidro. This choice also makes sense because, when referring to Ferdinando's ego-network shown in Figure 3, the only node connected to two still unidentified nodes is Node 1. This fits my opinion that David Vidro would naturally be in contact with his two brothers, Juan and Jorge. With this assumption, I assume nodes 2 and 3 to be either Juan or Jorge. There was no further evidence to determine which was which. For clarity, I present Ferdinando's ego-network in Figure 5 which the newly labeled discoveries.

Now that the Catalano/Vidro social structure has been defined, we can examine how it changes over the course of the 10 days.

3 THE CHANGING NETWORK

In order to view how the structure changes over time, we once again use the timeline view of *SocialAction*. In Figure 6, I present a simplified timeline, using only the nodes in the Paraiso subnetwork of Figure 4.

From this visualization, it is evident that the Vidros and Catalanos were very active until June 8. One June 8, all 5 individuals are

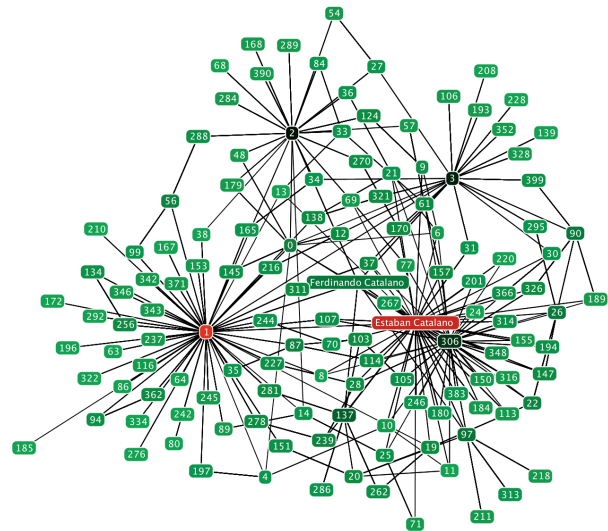


Figure 4. Approximating the subnetwork of the Paraiso movement. This is the subnetwork of Ferdinando Catalano (Node 200), his connections, and his connection's connections. The nodes are colored according to their degree.

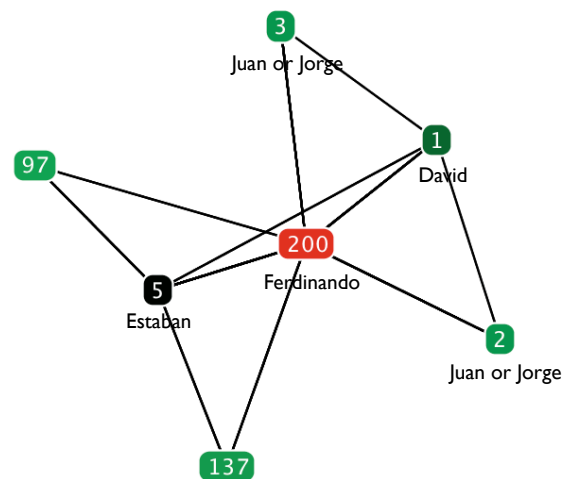


Figure 5. Updating the ego-network of the node assumed to be Ferdinando Catalano (Node 200) with the other assumptions of Estaban Catalano and the Vidro brothers.

very quiet, with more activity on June 9, and even more on June 10. However, on June 10, each of them is taking part in phone calls. In the bottom right portion of the visualization, it is clear that an individual (Node 306) becomes very active starting on June 8.

SocialAction's TimeLine is interactive, as demonstrated in the video. In Figure 7, I illustrate this by showing the TimeLine with only the Vidro and Catalanos.

This illustrates the change in activity even more dramatically among the identified individuals. This also makes it quite clear that David Vidro is probably no longer coordinating Paraiso activities on a daily basis over the phone, as his degree in the later days is minimal.

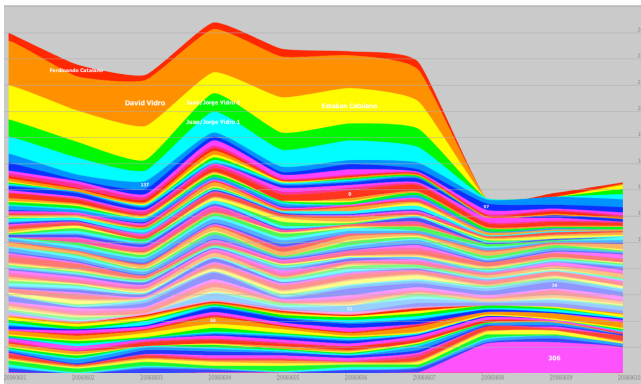


Figure 6. The TimeLine visualization illustrates the change in activity of the nodes in the Paraiso subnetwork displayed in Figure 4.

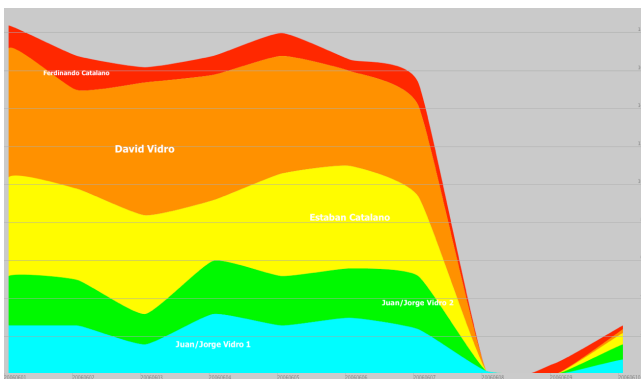


Figure 7. The TimeLine visualization illustrates the change in activity of the nodes in Fernando's ego-network displayed in Figure 3.

Now that we have an overview, it is also useful to coordinate this view with the actual network visualizations. SocialAction allows you to view networks at specific points in time. A collection of visualizations from each day is displayed at the end of this document. Each of these network visualization's nodes have been ranked according to their degree centrality, so the most-connected nodes are red and less-connected nodes are green. This provides consistent images with the Timeline visualization, as Fernando, David, Juan, Jorge, and Estaban are all very active until June 8. The June 8 network visualization shows a very disconnected network, except for node 308. The communication levels of

Fernando, David, Juan, Jorge and Estaban do increase and they are all part of phone calls on June 10. However, none of these individuals are talking to each other.

The combination of the TimeLine visualization and the daily network visualizations provide insight into how the social structure of the Catalanos and Vidros changed over the course of 10 days.

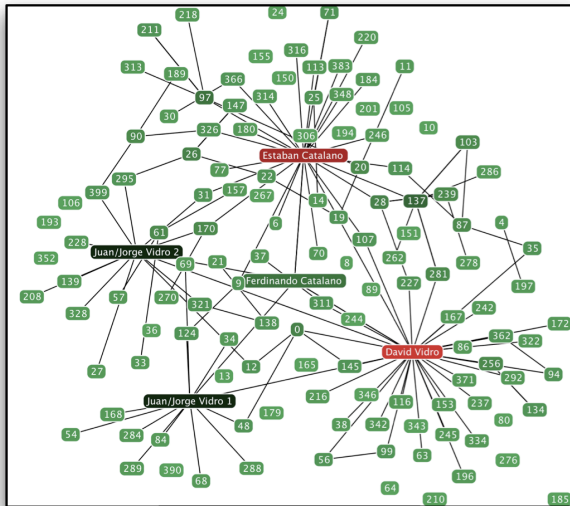
4 CONCLUSION

This document illustrates how SocialAction was used to make discoveries in the VAST Challenge data sets. Although the data consisted of anonymized caller records and a few clues, some interesting discoveries were made using social network analysis and visual analytics. Although I can not be 100% certain of my claims, they provide interesting hypotheses for further investigation. For instance, why did the activity of the Catalano's and Vidro's decrease after June 8. Who is Node 306 that emerged during this quiet time? Any why, after the Catalanos and Vidros reemerge on their cell phones do they not contact each other. Admittedly, the one piece of data I ignored was the map coordinate data, as I didn't see how it fit into the analysis. However, its possible additional clues were missed due to my judgment on this issue.

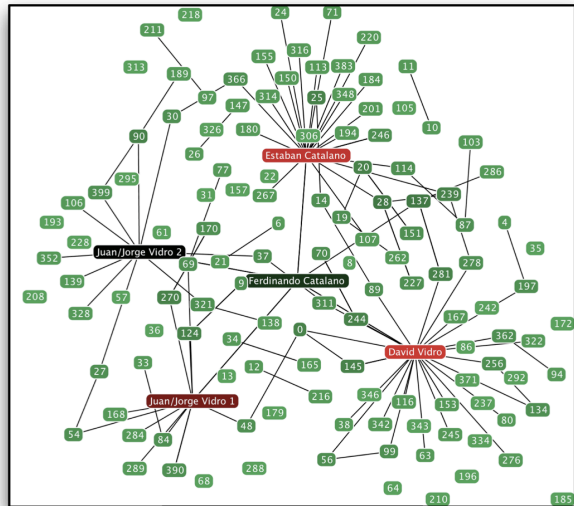
Although I may not have found every hidden pattern in the data set, I was able to answer each of the questions I asked the tool. SocialAction, which was designed to support exploratory data analysis, allowed me to make discoveries, highlighting the power of visual analytic tools. Examples of how SocialAction was used by real analysts over the course of many weeks are shown in .

REFERENCES

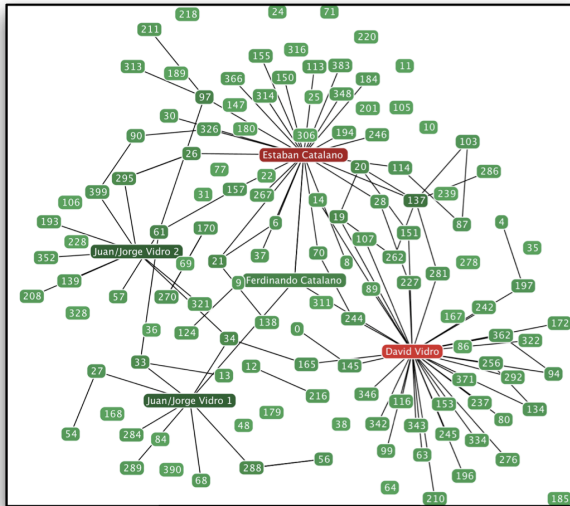
- [1] Adam Perer, Ben Shneiderman: Balancing Systematic and Flexible Exploration of Social Networks. IEEE Transactions on Visualization and Computer Graphics (InfoVis 2006). 12(5): 693-700. Baltimore, United States. (2006).
- [2] Adam Perer and Ben Shneiderman. Integrating Statistics and Visualization: Case Studies of Gaining Clarity During Exploratory Data Analysis. (CHI 2008). Florence, Italy. (2008).
- [3] Wattenberg, Martin. Baby Names, Visualization, and Social Data Analysis. IEEE Symposium on Information Visualization. (InfoVis 2005). Minneapolis, Minnesota, USA. (2005).



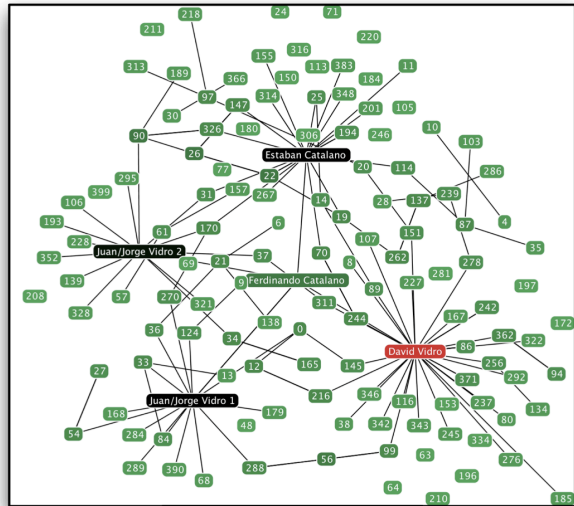
June 1, 2006



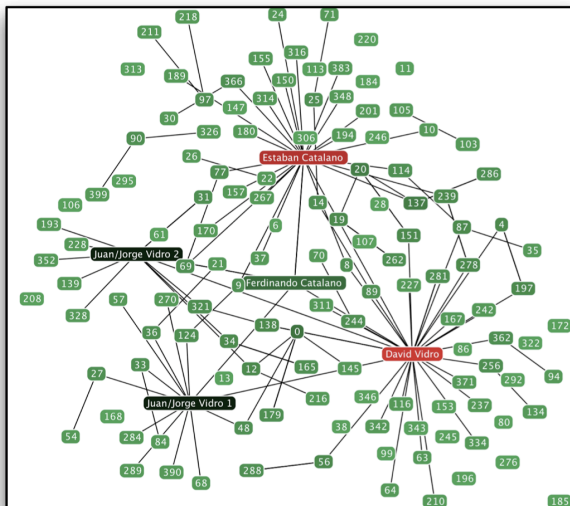
June 2, 2006



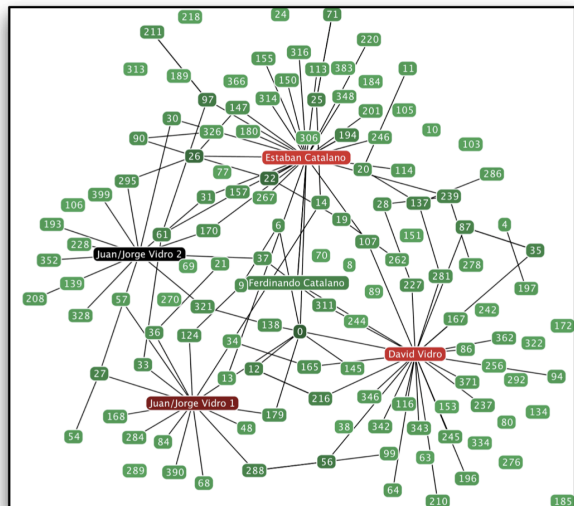
June 3, 2006



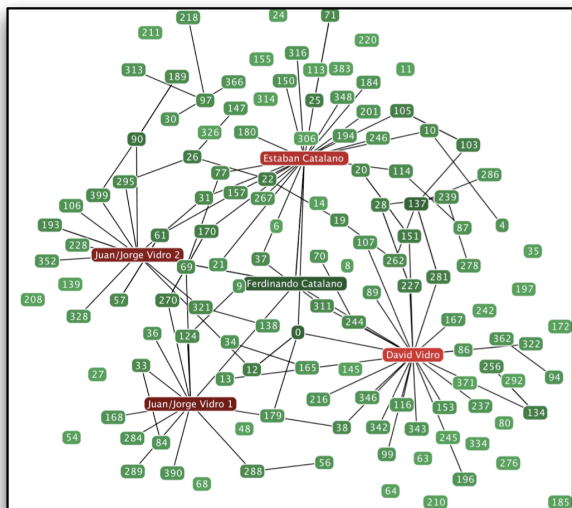
June 4, 2006



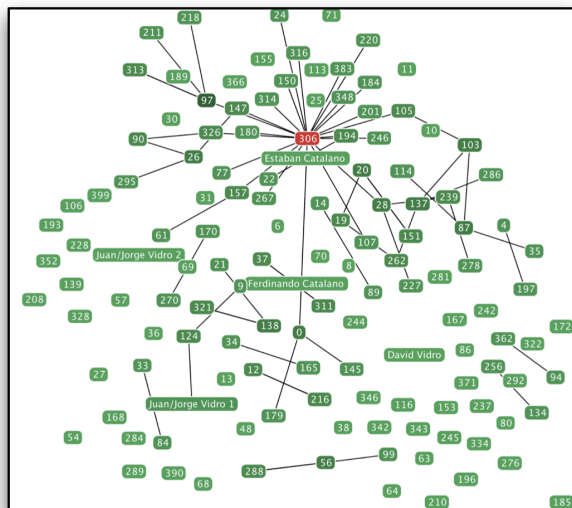
June 5, 2006



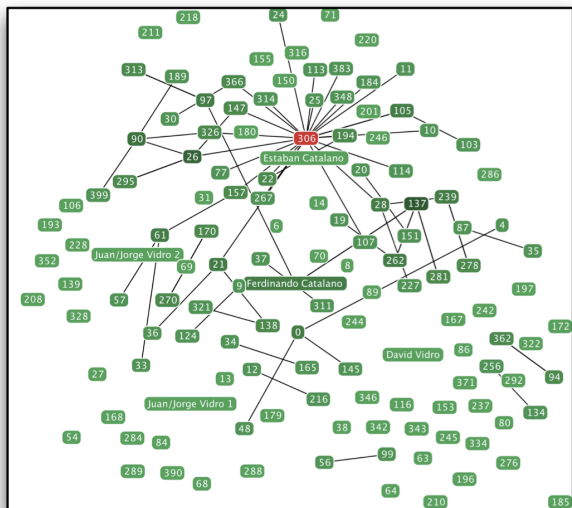
June 6, 2006



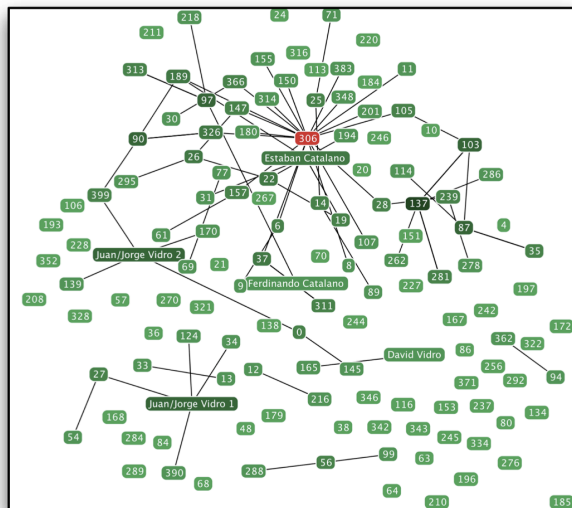
June 7, 2006



June 8, 2006



June 9, 2006



June 10, 2006